



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery

**Citation for published version:**

Laing, R, Kikuchi, T, Martinelli, A, Tsai, I, Beech, R, Redman, E, Holroyd, R, Bartley, D, Beasley, H, Britton, C, Curran, D, Devaney, E, Gilabert, A, Hunt, M, Johnston, S, Kryukov, I, Li, K, Morrison, A, Reid, A, Sargison, N, Saunders, G, Wasmuth, J, Wolstenholm, A, Berriman, M, Gilleard, J & Cotton, J 2013, 'The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery', *Genome Biology*, vol. 14, no. 8, R88. <https://doi.org/10.1186/gb-2013-14-8-r88>

**Digital Object Identifier (DOI):**

[10.1186/gb-2013-14-8-r88](https://doi.org/10.1186/gb-2013-14-8-r88)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Genome Biology

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## Title

The genome and transcriptome of *Haemonchus contortus*: a key model parasite for drug and vaccine discovery

## Authors

Roz Laing<sup>a</sup>, Taisei Kikuchi<sup>b,c</sup>, Axel Martinelli<sup>b</sup>, Isheng J. Tsai<sup>b,c</sup>, Robin N. Beech<sup>d</sup>, Elizabeth Redman<sup>e</sup>, Nancy Holroyd<sup>b</sup>, David J. Bartley<sup>f</sup>, Helen Beasley<sup>b</sup>, Collette Britton<sup>a</sup>, David Curran<sup>g</sup>, Eileen Devaney<sup>a</sup>, Aude Gilabert<sup>g</sup>, Martin Hunt<sup>b</sup>, Frank Jackson<sup>f</sup>, Stephanie Johnston<sup>a</sup>, Ivan Kryukov<sup>g</sup>, Keyu Li<sup>g</sup>, Alison A. Morrison<sup>f</sup>, Adam J. Reid<sup>b</sup>, Neil Sargison<sup>h</sup>, Gary Saunders<sup>a,b</sup>, James D. Wasmuth<sup>g</sup>, Adrian Wolstenholme<sup>i</sup>, Matthew Berriman<sup>b</sup>, John S. Gilleard<sup>e\*</sup>, James A. Cotton<sup>b\*</sup>

\*corresponding authors

## Affiliations

a Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

b Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

c Division of Parasitology, Department of Infectious Disease, Faculty of Medicine, University of Miyazaki, Miyazaki, 889-1692 Japan

d Institute of Parasitology, McGill University, Ste Anne de Bellevue, Quebec, Canada

e Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

f Disease Control, Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, United Kingdom

g Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

h Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom

i Department of Infectious Diseases and Center for Tropical and Emerging Global Disease, University of Georgia, Athens, USA

\*JSG: +1 403 210 6327, jsgillea@ucalgary.ca; JAC: +44 1223 494864, james.cotton@sanger.ac.uk

Email addresses. RL: Rosalind.Laing@glasgow.ac.uk, TK: tk6@sanger.ac.uk, AM:

am21@sanger.ac.uk, IJT: jit@sanger.ac.uk, RNB: robin.beech@mcgill.ca, ER:

libbyredman@hotmail.co.uk, NH: neh@sanger.ac.uk, DJB: Dave.Bartley@moredun.ac.uk, HB:

hr1@sanger.ac.uk, CB: Collette.Britton@glasgow.ac.uk, DC: dmcurren@ucalgary.ca, ED:

Eileen.Devaney@glasgow.ac.uk, AG: agilaber@ucalgary.ca, MH: mh12@sanger.ac.uk, FJ:

Frank.Jackson@moredun.ac.uk, SJ: s.johnston.2@research.gla.ac.uk, IK: ikryukov@ucalgary.ca, KL:

keli@ucalgary.ca, AAM: Alison.Morrison@moredun.ac.uk, AJR: ar11@sanger.ac.uk, NS:

Neil.Sargison@ed.ac.uk, GS: gary.saunders@sanger.ac.uk, JDW: jwasmuth@ucalgary.ca, AW:

adrianw@uga.edu, MB: mb4@sanger.ac.uk, JSG: jsgillea@ucalgary.ca, JAC:

james.cotton@sanger.ac.uk

## Abstract

**Background:** The small ruminant parasite *Haemonchus contortus* is the most widely used parasitic nematode in drug discovery, vaccine development and anthelmintic resistance research. Its remarkable propensity to develop resistance threatens the viability of the sheep industry in many regions of the world and provides a cautionary example of the effect of mass drug administration to control parasitic nematodes. Its phylogenetic position makes it particularly well placed for comparison with the free-living nematode *Caenorhabditis elegans* and the most economically important parasites of livestock, as well as the human hookworms.

**Results:** Here we report the detailed analysis of a draft genome assembly and extensive transcriptomic dataset for *H. contortus*. This represents the first genome to be published for a strongylid nematode and the most extensive transcriptomic dataset for any parasitic nematode reported to date. We show a general pattern of conservation of genome structure and gene content between *H. contortus* and *C. elegans*, but also a dramatic expansion of important parasite gene families. We identify genes involved in parasite-specific pathways such as blood feeding, neurological function, and drug metabolism. In particular, we describe complete gene repertoires for known drug target families, providing the most comprehensive understanding yet of the action of several important anthelmintics. Also, we identify a set of genes enriched in the parasitic stages of the lifecycle and the parasite gut that provide a rich source of vaccine and drug target candidates.

**Conclusions:** The *H. contortus* genome and transcriptome provides an essential platform for post-genomic research in this and other important strongylid parasites.

## Keywords

nematode, parasite, anthelmintic, resistance, genome, transcriptome

## Background

Resistance to broad spectrum anthelmintic drugs is now widespread in parasites of domestic livestock [1, 2] and there are increasing concerns about the sustainability of human parasite control programmes using mass administration of the same drugs [3]. Consequently, there is an urgent need to understand the genetic mechanisms underlying anthelmintic resistance and to discover new methods of chemical and non-chemical control. However, the genomic and genetic resources required to underpin research in parasitic nematodes are lacking [4-6]. The free-living nematode *Caenorhabditis elegans* is a powerful model system, but it has clear limitations for the study of parasitic species [7]. Although the need to develop workable parasitic nematode model systems is widely recognised, most human helminth species are not amenable to experimental study. In contrast, *Haemonchus contortus*, a gastrointestinal parasitic nematode of small ruminants, has a successful track record in anthelmintic resistance [7, 8], drug discovery [9, 10] and vaccine [11-13] research. It is amongst the most experimentally tractable parasites for a number of reasons: adult females are relatively large and produce thousands of eggs per day allowing the production of large amounts of biological and genetic material, the infective larvae (L3) can be viably cryopreserved, and *in vivo* studies, including genetic crosses, can be undertaken in the natural host [8]. Its phylogenetic position within the most closely-related group of parasites to *C. elegans* facilitates comparative genomics and heterologous gene expression allowing functional studies to be performed on *H. contortus* genes and regulatory elements [4, 14]. As it is a strongylid nematode, research on this parasite is of particular relevance to the most economically significant parasites of grazing ruminants and to the human hookworms [15]. *H. contortus* itself causes major economic loss in small ruminants worldwide; it is highly pathogenic and unsurpassed in its ability to develop resistance to every anthelmintic used in its control. Notably, these include a number of core drugs used for mass drug administration programmes in humans [3]. Anthelmintic resistance in *H. contortus* and related strongylids now threaten the viability of the sheep industry throughout the world [2]. This

represents both a warning and a useful model for the consequences of the widespread intensive use of anthelmintics that are now being used to control human parasites in the developing world [3].

*H. contortus* has a direct and rapid lifecycle (Figure S1 in Additional file 1); adults reside and mate in the abomasum of the ruminant host, then females produce eggs to be excreted in the faeces. The eggs embryonate, develop and hatch as first stage larvae (L1), develop and moult to become second stage larvae (L2), then moult again to become third stage larvae (L3) in the faeces. The L3 migrate onto pasture to be ingested by the grazing ruminant host. The L3 shed the retained L2 cuticle (exsheath) in the rumen, travel to the abomasum and develop through to the fourth larval stage (L4) then adults in two to three weeks [16]. As voracious blood-feeders, *H. contortus* L4 and adults can cause severe haemorrhagic gastritis, hypoproteinaemia, anaemia and oedema, with acute infections resulting in death of the animal host. One adult female can produce up to 10,000 eggs per day [17] and a single animal can harbour thousands of worms. This extremely high fecundity, under conducive host and environmental conditions, can give rise to population explosions and devastating disease outbreaks. Development occurs most rapidly in warm humid conditions, but the L4 stage can undergo arrested development within the host to survive adverse conditions such as prolonged drought or cold winters [18]. This feature of facultative arrested development, aided by movement of domestic livestock and climate change, has ensured the worldwide distribution and success of this parasite even though its evolutionary origins were in sub-Saharan Africa [19].

One striking feature of *H. contortus*, in common with related nematode species, is the extremely high level of genetic diversity that has been reported in both laboratory and field populations; this is thought to be predominantly a function of its large effective population size [20-22]. Genetic variation may underlie both the parasite's remarkable ability to adapt to different climatic regions and host species [23] and its alarming propensity to develop drug resistance. This high level of genetic polymorphism has also provided a major challenge to genome assembly necessitating the production of an inbred line from which to prepare DNA template.

In this paper, we describe the assembly and annotation of the draft genome of MHco3(ISE).N1, a genetically well characterised inbred *H. contortus* strain that is susceptible to all major anthelmintic drugs. The most extensive comparative transcriptomic sequencing and analysis yet described for a parasitic nematode was undertaken on representative life-stages and the nematode gut to explore different aspects of nematode biology, evolution and parasitism. This draft genome provides a platform for future research in anthelmintic resistance, drug discovery and vaccine development using *H. contortus*, currently the most important experimental model for the strongylid nematode group.

## Results and Discussion

### Genome structure and content

We assembled a draft sequence of the *H. contortus* genome based on data from a mixture of sequencing technologies (see Material and Methods, and Supplementary results and methods in Additional file 1). Our final draft assembly consists of 67,687 contigs linked into 26,044 scaffolds of total length 370 Mb, including 23.8 Mb of inferred gaps between contigs, with an average and N50 scaffold length of 14,206 and 83,287 bp respectively (Table S1 in Additional file 1). This is a significantly larger genome size than the ~60Mb predicted with flow cytometry [24] but it is consistent with a prediction of ~200-300Mb inferred from a pilot annotation of two large contiguous regions of the genome [14], making this the largest nematode genome sequenced to date (Table S1 in Additional file 1). The overall base composition of all sequence contigs was close to neutral at a mean 41.3% GC, slightly higher than that for other clade V nematodes. Approximately 93% of conserved eukaryotic genes can be identified in the assembly, suggesting that our draft assembly represents at least that fraction of the *H. contortus* genome [25], as many of these models are incomplete or split between contigs (Table S2 in Additional file 1). Despite this, our current assembly is of similar quality to some other published draft nematode genome sequences [26-30] (Table S1 in Additional file 1).

We used extensive transcriptomic evidence from across the *H. contortus* life cycle (see Methods and Table S3 in Additional file 1) to guide *de-novo* gene model curation for protein-coding genes. We predict a similar number of protein-coding genes to the *C. elegans* genome (21,799 vs. 20,532), but significantly lower gene density of 59 genes per Mb, with only seven percent of the genome being protein coding, compared to 200 genes per Mb and a protein coding content of ~30% in the *C. elegans* genome (Figure 1). The average coding sequence length is similar in *H. contortus* and *C.*



*elegans* (1127 bp in *H. contortus* compared to 1371 bp in *C. elegans*) but the average gene length is more than double in the parasite (6564 bp compared to 3010 bp in *C. elegans*). A closer look at a subset of 2822 *H. contortus* and *C. elegans* one-to-one orthologs shows the discrepancy is explained by an expansion in both the number and length of introns in *H. contortus* (average of ten introns per gene, average size 633 bp; relative to six introns per gene, average size 340 bp in *C. elegans*). An expansion in intronic sequence has also been reported in the closely related necromenic nematode *Pristionchus pacificus* but to a less dramatic extent (average of nine introns per gene, average size 309 bp) [26]. Over 80% of *H. contortus* and *C. elegans* one-to-one orthologs present on the same scaffold occur on the same *C. elegans* chromosome (Table S4 in Additional file 1), suggesting widespread conservation of synteny between the two species (Figure 2). However, gene order is generally poorly conserved, which is consistent with comparisons between *C. elegans* and related nematodes [27, 31], but regions of conserved microsynteny are apparent and may prove useful in supporting orthology of divergent genes.

Reflecting this relatively expanded genome, approximately 29% of the draft genome assembly consists of repetitive sequence (compared to ~16% in *C. elegans*) including 2434 copies of the characteristic HcRep element previously reported from a number of trichostrongylid species [14, 32-34]. The repetitive sequence also includes representatives of many known repeat families in other nematodes, with approximately 6% of the genome derived from LINEs, 2.5% from long terminal repeat retrotransposons and 5% from DNA transposons, including TTAA-specific, Mariner-like and MuDR-superfamily elements, together with evidence of elements related to Tc1 and Tc4 of *C. elegans*. Despite belonging to similar families, *H. contortus* repeats represent independent expansions to those in other clade V nematodes, as repeat libraries from *H. contortus* show little similarity to genome sequence from other species, and vice-versa. Our transcriptomic data suggests that active transposition is occurring, with evidence of expression for four out of 26 gene models annotated with transposase domains and 49 out of 482 reverse transcriptase domain-containing

proteins. Expressed proteins appear to belong to a range of DNA transposon types, and to Gypsy-related and LINE retro-elements.

In *C. elegans*, around 17% of genes are in operons [35]: tightly linked clusters of two-to-eight genes, which are co-transcribed from the same promoter. The resulting polycistronic pre-mRNAs are resolved by *trans*-splicing with spliced leader (SL) SL1 and SL2 sequences. Most frequently, SL1 is *trans*-spliced to the first gene in an operon and downstream genes are SL2 *trans*-spliced. The structure (gene complement, order and orientation) of around 23% of *C. elegans* operons is conserved in the *H. contortus* genome. The structure of a further 10% of *C. elegans* operons appear to be partially conserved, where at least two orthologs are present on the same scaffold in the expected order and orientation, but one or more genes are in a different order, inverted or absent. Functional constraints are thought to conserve the intergenic distance in *C. elegans* operons to ~100 bp but genes in *H. contortus* operons are further apart: the average intergenic distance of genes with a conserved operon structure is 992 bp (median 621 bp, largest 8329 bp), and the operon encoding ion channel subunits Hco-des-2H and Hco-deg-3H has an intergenic distance of 2342 bp [36].

Overall, SL1 *trans*-splicing was detected in 6306 *H. contortus* genes and SL2 *trans*-splicing was detected in 578 genes. Of these, 318 *trans*-spliced genes were in the putative conserved operons identified above (see Additional methods in Additional file 1). All 126 first genes in operons were *trans*-spliced to SL1 (SL2 *trans*-splicing was detected in five putative first genes, but examination of their loci suggests they are downstream genes in new operons in this species); 119 downstream genes were *trans*-spliced to SL1 and 73 were *trans*-spliced to SL2. If SL2 *trans*-splicing is the definitive criterion in identifying operons, the relatively low level of SL2 *trans*-splicing to downstream genes suggests that either operon function is less frequently conserved than operon structure in *H. contortus* or that undetected, divergent SL2 sequences are present. However, the relatively high frequency of SL1 *trans*-splicing in genes that are also SL2 *trans*-spliced (~77%; 56 of 73 in conserved

operons, 445 of 578 in all genes identified) suggests SL1 *trans*-splicing of downstream genes may also be relatively common in this species.

We employed two complementary approaches to global comparison of the *H. contortus* gene set, using the Inparanoid algorithm to look in detail at orthologs with *C. elegans* and *Pristionchus pacificus*, and OrthoMCL for a wider view of gene family evolution with other clade V nematodes. Of 5,937 orthology groups between *C. elegans* and *H. contortus*, 5,012 are one-to-one orthologs, while an additional 899 orthologs could be identified in *H. contortus* and *P. pacificus* but not *C. elegans*, suggesting they have been lost in the *C. elegans* lineage (Table S5 in Additional file 1). A number of orthology groups are significantly expanded in *H. contortus*, including a family of 180 *Haemonchus* paralogs to a single *C. elegans* gene that lacks any functional annotation (Table S6 in Additional file 1). Other expanded groups include genes with likely roles in parasitism, such as cysteine-rich secreted proteins, together with a set of helicase domains that include some with predicted signal peptides. Global analysis of the evolution of entire gene families (clusters of similar genes) across the clade V nematodes confirms this pattern of significant diversity within the clade (Figure 1), and allowed us to identify *H. contortus* genes lacking clear orthologs in other clade IV or V nematodes (Figure S2, Table S8 in additional file 1). This shows that the *Haemonchus*-specific proteome is enriched in genes encoding polypeptides involved in proteolysis, neurotransmission and carbohydrate metabolism, and in secreted proteins and those expressed in the cuticle. While some of these genes are explored in detail in the more focused analyses below, others may represent novel candidate genes involved in the host-parasite interface.

### **Gene Expression in Parasite Life-stages and Gut**

As *H. contortus* progresses through its life cycle, it must adapt to different environments with differing food sources and energy requirements (see Figure S1 in Additional file 1) and this is reflected in differential gene expression. RNA-seq was used to analyse gene expression in six parasite life-stages and the adult female gut. Samples were made in triplicate from independent

batches of parasite material for every stage allowing statistically robust comparison of relative gene expression between the stages (see Methods and Table S3 in Additional file 1). We found significant expression for 17,483 genes in total from the six life stages studied; with between 13,962 (in L3) and 15,569 (in adult males) genes expressed in each stage. A total of 11,295 genes were significantly up or down-regulated through the life-cycle (see Figure 3 and Materials and Methods) and we used annotation with Gene Ontology (GO) terms to investigate their broad functions. Metabolic enzyme expression throughout the parasite lifecycle was analysed in more detail and will be discussed separately.

Genes up-regulated in the development from egg to the L1 stage include those associated with muscle development and motor activity, while up-regulated in the egg are genes associated with oxidoreductase activity, apoptosis, body morphogenesis and development, larval and embryo development, as well as DNA replication and chromosome organisation. This pattern is consistent with the progression from a developing embryonated egg to a motile and actively feeding larval stage.

In the transition to the L3 stage, a decrease in the expression of genes associated with the myosin complex and motor activity and various metabolic processes are consistent with the nematode entering a quiescent state. Among the up-regulated genes there is an association with oxygen transport and heme binding. Oxidoreductase enzymes are over-represented and may reflect the increased need to detoxify a build up of endogenous waste, consistent with previous studies showing higher cytochrome P450 (CYP) activity in the *H. contortus* L3 than in L1 or adult stages [37]. CYPs have also been shown to be up-regulated in response to reduced food intake [38]. This is consistent with the significant increase in gluconeogenesis from the L1 to L3, which is also increased in the *C. elegans* dauer [39], and the up-regulation of acetyl-CoA metabolic process, likely to reflect metabolism of fat stores in the non-feeding L3. Genes associated with binding of cobalamin (Vitamin B12) are also up-regulated. Cobalamin has been shown to be strongly concentrated and stored in

the infective L3 of other gastrointestinal nematodes [40] and a ready supply may be required for rapid larval development after ingestion by the host.

The L4 is the first blood-feeding stage of *H. contortus*. The transition from the quiescent L3 stage to the motile L4 stage is reflected in significant up-regulation of many genes, including genes associated with motor activity, the myosin complex and locomotion as well as various metabolic processes. The binding of oxygen, lipids and sugars, possibly associated with active feeding, are also up-regulated as are changes in the expression of genes linked to response to oxidative stress that may reflect the reactivation of the parasite from its dormant stage. A significant increase in the expression of genes associated with collagen and cuticle development and body morphogenesis, consistent with parasite growth, is also observed. Interestingly, heme-binding genes were both up- and down-regulated, perhaps reflecting an increase in heme load from blood feeding and a decrease in CYP activity.

Transition from the L4 to the adult stages is characterised by several changes. In the transition to the female stage, 1658 genes are up-regulated and correlate with gender-specific development as well as embryogenesis (adult females contain oocytes and eggs at various developmental stages), such as genitalia development, embryo development, oogenesis, ovulation, germ-line cell cycle switching, meiosis regulation and regulation of vulval development. A significant increase in DNA replication processes is also apparent. Between the L4 and adult male stage, lower expression in the adult male of genes linked with body morphogenesis, molting, collagen and cuticle development, oxidoreductase activity, heme binding and response to oxidative stress were observed among the various alterations in expression. Increased expression of genes annotated with the GO term 'structure molecule activity' is due to up-regulation of a number of major sperm protein genes. These sex-specific gene expression patterns were confirmed by comparison between adult female and male parasites, and it is clear that many genes are highly enriched in male worms, being expressed only at low levels in other stages. The large set of genes highly expressed in both eggs and adult females were again apparent.

Finally, we investigated gene expression in the *H. contortus* intestine, the major organ of digestion and detoxification in the nematode, by comparing the female gut sample with the whole female worm. Consistent with data from *H. contortus* gut EST libraries [41], increased expression of genes with protein kinase, cysteine-type peptidase and cysteine-type peptidase inhibitor activities predominated. Genes associated with sugar and cobalamin binding were significantly up-regulated as were genes associated with transport of cations, anions and oligopeptides. Oxidoreductase activity was also increased, consistent with the expression pattern of detoxification genes in *C. elegans* [42].

### **Metabolic Pathways and Chokepoint Analysis**

Comparisons between *H. contortus* and the free-living nematodes revealed 22 enzyme classifications (ECs) that were restricted to the parasite (Table S9 in Additional file 1). While more detailed analysis is required, metabolism of amino acids and carbohydrates clearly differ between these two groups. For example, lysine 6-aminotransferase (EC 2.6.1.36) catalyses lysine to glutamate, which can be further converted to  $\alpha$ -ketoglutarate, an intermediate of the tricarboxylic acid (TCA) cycle. Lysine 6-aminotransferase was previously considered restricted to prokaryotes, thus its activity in *H. contortus* needs to be confirmed.

A summary of up- and down-regulated metabolic enzymes across all life-stages is shown in Table S10 (in Additional file 1). The transition through eggs, L1, L3 and L4, showed a striking pattern: from L1 to L3, most enzyme classifications were down-regulated, including those involved in carbohydrate, lipid and energy metabolism, but many of these are up-regulated again in the transition to L4. This is consistent with the L3 being a stage in which development is arrested analogous to the dauer larva in *C. elegans*. Further support for this comparison is the up-regulation of two enzymes that independently convert isocitrate to 2-oxo-glutarate, while most other parts of the TCA cycle are down-regulated. Furthermore, 2-oxo-glutarate is an entry metabolite into the ascorbate and

aldarate metabolic pathway, which is implicated in increased lifespan in *Drosophila* [43]. The L4-to-male transition shows a decrease in lipid metabolism coupled with an increase in amino acid metabolism.

Metabolic chokepoints – reactions that uniquely consume or produce a metabolite – are enzymes that seem likely to be essential to the parasite and so may be potential targets for future drug development. Analysis of the *H. contortus* network revealed 362 chokepoint reactions, five of which passed additional criteria of lacking isoenzymes and being divergent from known human enzymes [44] (see Table S11 in Additional file 1). Comparison of these against the Therapeutic Target Database (TDD) revealed that two are known potential targets for anthelmintic drugs: trehalose-6-phosphatase (EC 3.1.3.12) is a member of the sucrose metabolism pathway, and is currently being researched as a potential drug target in the filarial nematode *B. malayi* [45], while dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13) is currently of interest in *Mycobacterium tuberculosis* [46]. This validation of the approach justifies further analysis of the other three chokepoint enzymes identified.

### **Neuromuscular Drug Targets**

In nematodes, the pentameric ligand-gated ion-channel (pLGIC) family is particularly numerous, with 64 members identified in *H. contortus* via homology with *Caenorhabditis* spp., *P. pacificus* and RNA-seq data (Figure 4). They are of great importance in parasitic nematodes because they are targets of the majority of the currently available anthelmintic drugs (as summarized in Table S12 in Additional file 1; the  $\beta$ -tubulin targets of the benzimidazole group have been described previously [47]).

The pLGICs regulate the flow of anions, typically chloride ions, or cations, including sodium and calcium, in response to an extracellular signal in the form of an activating ligand or change in pH. They are fundamental to synaptic transmission; interference with their normal function results in paralysis and death. Drugs that activate the anionic channels, such as the macrocyclic lactone

ivermectin (IVM), typically inhibit neuronal transmission and muscle contraction. Those that activate the cationic channels, such as levamisole (LEV) and monepantal, stimulate neuronal transmission and typically induce muscle contraction. Here we present the most complete picture of these channels to date and show that as expected, this parasite is very similar to *C. elegans* and supports the use of the free-living worm as a functional model for the parasite nervous system. There are however some important differences, most significantly in glutamate signaling that is sensitive to the macrocyclic lactones and acetylcholine signaling that sensitive to levamisole, as well as characteristic loss of some receptor genes associated with biogenic amine signaling.

The macrocyclic lactones, which include IVM and moxidectin, act at several different glutamate-gated chloride channels. Some of these are found in both *C. elegans* and *H. contortus*, but it is noteworthy that two *H. contortus* subunit genes, *glc-5* and *glc-6*, encode glutamate sensitive channels that are absent from *C. elegans*. It is likely these were lost from the rhabditid lineage as homologous sequences can be detected in the genome of the close relative *P. pacificus*. Both of these subunits are targets for the macrocyclic lactones [48, 49] and changes in either their sequence or expression have been associated with drug resistance in veterinary parasites [50]. The archetypal target of IVM, *Cel-glc-1*, appears to be a duplication of *avr-15*, specific to *C. elegans*. These differences may explain the difficulties encountered in understanding ivermectin's mode of action and resistance in parasitic nematodes by studying the model organism [51]. Most of the other anionic channel subunits in *H. contortus* have direct orthologs in *C. elegans* (Figure 4A). The only notable absences are *lgc-48* and *lgc-54*. *lgc-48* encodes an as yet uncharacterized member of the *acc-1* acetylcholine gated chloride channels [52] that appears to have been lost from *H. contortus*, and *lgc-54* encodes an uncharacterized member of the *C. elegans* biogenic amine-gated chloride channel family that may not be present in *H. contortus* [53].

The acetylcholine gated receptor (AChR) cation channels are of particular interest as they are the targets of several anthelmintic drugs already in use (levamisole, monepantel and derquantel) and



are considered to be one of most promising gene families for the identification of new drug targets. Comparison of the *H. contortus* gene family with those found in *C. elegans* is shown in Figure 4B. There is a one-to-one correspondence in *H. contortus* for subunits of the *acr-16* clade [54], named for the homomeric nicotine sensitive channel in *C. elegans*, and *acr-16* expression is much higher in adult males than in other life-stages, as previously described for *C. elegans* [55]. The anthelmintic levamisole targets receptors composed of  $\alpha$  subunits of the *unc-38* clade and non- $\alpha$  subunits from the *unc-29* clade. In *C. elegans* three different alpha and two different non-alpha subunits combine to form a channel that responds to acetylcholine, strongly to levamisole and only weakly to nicotine [56]. The  $\alpha$ -type *acr-13* (*lev-8*) gene, required in *C. elegans* for the LEV receptor, appears to have been lost in *H. contortus* as an ortholog of this gene is detectable in *P. pacificus*. The non-alpha gene *lev-1* is present, but the signal peptide appears to have been lost [57]. A LEV receptor in *H. contortus* can be reconstituted without either of these, requiring only UNC-38, UNC-63 and UNC-29 as in *C. elegans* [57] and replacing ACR-13 with ACR-8. Most strikingly, the non- $\alpha$  subunit gene *unc-29* in *C. elegans* corresponds to four paralogous copies in *H. contortus* [57]. A LEV sensitive channel has been reconstituted containing the Hco-UNC-29.1 subunit [58]. The degree to which the other copies may have diverged in function is currently under investigation. The *deg-3* clade [54] encodes subunits that form channels involved in chemotaxis, and is of particular interest as channels encoded by *des-2/deg-3* and *acr-23* in *C. elegans* are targeted by the relatively new anthelmintic monepantel (MPTL), with *acr-23* as the principle target of MPTL action *in vivo* [9]. The single subunit gene *mptl-1* in *H. contortus* corresponds to the *acr-23/acr-20* pair in *C. elegans*. Splice site mutations in the *mptl-1* gene are associated with resistance to MPTL [35], which would suggest that *acr-23* is functionally equivalent to *mptl-1*.

*H. contortus* possesses two AChR genes that are not present in *C. elegans*, *acr-26* and *acr-27*, and these two genes seem to form a distinct clade, which suggests that they are likely to form receptors with a distinct pharmacology. Orthologs of *acr-26* are found in many species of parasitic nematode

and therefore they might make interesting targets for the development of novel cholinergic anthelmintics.

In summary, the majority of the pLGIC subunit genes in *C. elegans* have direct orthologs in *H. contortus*, although there is more variation in the repertoire of orphan family pLGICs (see Additional results, Additional file 1). This supports the use of *C. elegans* as a model to study the neuromusculature of *H. contortus* in drug screens, but there are important differences in specific anthelmintic targets: for each of the anthelmintics ivermectin, levamisole and monepantel, the drug targets in *H. contortus* are functionally different to the *C. elegans* model.

### **Drug metabolism and efflux**

Parasitic nematodes are armed with a large repertoire of inducible metabolising enzymes and transporters to protect against environmental toxins. The nematode detoxification pathway can be divided into three main phases: modification, conjugation and excretion [59] involving the cytochrome P450s (CYPs) and the short-chain dehydrogenase/reductases (SDRs) in phase 1, the UDP-glucuronosyl transferases (UGTs) and the glutathione S-transferases (GSTs) in phase 2 and the ATP-binding cassette (ABC) transporters in phase 3. Little is known about the impact of the parasite detoxification system on anthelmintic efficacy or resistance [60], but a better understanding of these pathways will allow their role in anthelmintic resistance to be assessed and should also be informative in the design of new drugs and synergistic agents [61]. We have annotated a large number of modification and conjugation genes in the current assembly, identifying a total of 42 CYPs, 44 SDRs, 34 UGTs and 28 GSTs (see Figure S5 in Additional file 1), but here we focus on the excretion transporters implicated in drug resistance.

Members of the ATP Binding Cassette (ABC) transporter family hydrolyze ATP and couple the energy released with the active transport of a wide range of compounds including small organic molecules, lipids, proteins and metal ions. They are an essential component of many biological processes and

are fundamental to the barrier between a nematode and its environment. These transporters consist of a basic structure of six transmembrane (TM) domains with an associated intracellular ATP binding motif. The functioning transporter complex requires two of these protein halves, but some members of this family are fusion proteins with 12 TM domains and two ATP binding motifs. We find numerous differences in ABC transporter genes between *C. elegans* and *H. contortus* (Additional results and Figure S4, Additional file 1), with a reduced complement of haf-transporters including, for example, the loss of *haf-6*, which is essential for efficient RNAi in *C. elegans* [62], a significant expansion of *ced-7*, which has an unknown function in amphid and phasmid sensory cells and an extensive change in the repertoire of multidrug resistance protein genes. The P-glycoprotein (pgp) transporters are of particular interest as they have been implicated in resistance of *H. contortus* to ivermectin and other anthelmintics [49, 63], and this family has undergone significant change compared to the free living *C. elegans*. In total there are 10 P-glycoprotein genes identified in the *H. contortus* genome and this full complement will now allow a more systematic analysis of the role of P-glycoproteins in resistance to ivermectin and other anthelmintics. A cluster of four *C. elegans* genes, *pgp-5*, *6*, *7* and *8*, are not found in *H. contortus*. *Cel-pgp-3* and *4* as well as *Cel-pgp-12*, *13* and *14* represent gene duplication events corresponding to single genes in the parasite. *Cel-pgp-9* corresponds to two paralogous copies in *H. contortus* and in addition, two genes not present in *C. elegans*, related to *pgp-3* and *pgp-11*, have been retained in *H. contortus*. These are named *pgp-16* and *pgp-17*, respectively. Changes in the sequence or expression of *pgp-1*, *pgp-2* and *pgp-9* have been reported for ivermectin-resistant vs. susceptible isolates of *H. contortus* and in *pgp-9* for resistant *Teladorsagia circumcincta* [50, 64, 65].

### **Protease vaccine candidates**

*H. contortus* is a voracious blood-feeder with even modest infections of 1000 worms generating losses of up to 50 ml blood per day [66]. Cysteine, aspartic and metallo-proteases as well as aminopeptidases have been implicated in important aspects of parasite function including haemoglobin digestion and

anticoagulant activity, and these enzymes are also important vaccine candidates. Vaccination with gut extracts enriched for these activities can confer up to 75% reduction in worm burden and 90% reduction in egg output [67]. Protection is thought to result from the ingestion of host antibodies by the parasite during blood-feeding, which bind to gut antigens and disrupt function. Female parasites appear to be more affected than males increasing the relative impact on egg output. Development of a commercial vaccine, however, requires identification and expression of specific proteases which, either singly or combined, induce protective immunity. Comparative genomics and transcriptome data can aid target selection by identifying potential functionally important proteases and those enriched in the gut of blood-feeding L4 and adult stages, suggesting a role in blood-feeding as well as accessibility to host antibodies.

Cathepsin B protease (*cbl*) genes are part of an ordered haemoglobin degradation pathway, functioning after aspartic proteases (APR) and upstream of metalloproteases (MEP) and aminopeptidases [68, 69]. Cathepsin B diversity may therefore be key in generating an array of substrates from ingested nutrients for efficient cleavage by downstream proteases and may be involved in the high blood digestion capacity of *H. contortus*. Indeed, *H. contortus* has a higher copy number (63 genes) of cathepsin B protease (*cbl*) genes than related free-living nematodes, representing 80% of all cathepsin cysteine protease genes in the genome (Figure 5A). This large expansion of the cathepsin B family has resulted in *H. contortus* showing a greater diversity of *cbl* genes than known in any other parasitic nematode. Uniquely, members of each *cbl* type are arranged in tandem in the *H. contortus* genome (Figure 5B) and have the same genomic structure, suggesting that diversity has arisen through recent gene duplication. Reflecting this, most CBLs encoded in the *H. contortus* genome form large monophyletic groups distinct from *C. elegans* [70], hookworm [71] or other strongylid CBLs, suggesting that duplication and divergence has occurred separately following speciation. It is possible that gene amplification has occurred as a mechanism of increasing overall CBL expression associated with the need for *H. contortus* to digest the huge quantities of host blood it takes in during feeding. The *cbl* genes show increased expression in L4 and

adult stages and many are significantly enriched in gut tissue identifying these as potentially important control targets. These include both novel and previously identified *cbl* genes (*AC-4*, *gcp-7* and *hmcp-2*) [72-74], while other *cbl* transcripts are significantly up-regulated in the L4 stage and in adult male worms, but not in the gut, suggesting a role in development and reproduction rather than feeding. The CBLs identified here contain a putative N-terminal signal peptide and several have been identified in adult worm excretory-secretory (ES) products [75]. It has also been proposed that sequence variation of Hc-CBLs may confer antigenic diversity [76], and presentation to the host immune system through secretion may therefore drive the maintenance of the diversity of *Hc-cbl* genes. 3D modelling of the CBL repertoire and epitope mapping will clarify this issue.

Other cathepsin cysteine protease genes are not expanded – for example, a *cpr-6*-like single copy gene is highly conserved in a number of parasitic and free-living nematodes, suggesting a house-keeping role, and there is no expansion found for cathepsin L, F or Z genes in *H. contortus*.

Furthermore, there is less expansion of genes encoding other vaccine candidates; H-gal-GP composed predominantly of pepsinogen-like aspartic and metallo proteases [77], and H11 aminopeptidases [78]. These are integral gut membrane proteins, considered hidden from the immune system during natural infection [67] and lower diversity should be advantageous in vaccination studies.

Significant expansion is, however, seen in the cathepsin aspartic protease family (Figure S6, Additional file 1). Importantly, the genome data identifies a novel, single-copy *apr* gene with 84% amino acid identity to APR-1 of the dog hookworm *Ancylostoma caninum* and which is significantly enriched in gut tissue. Vaccination with recombinant Ac-APR-1 significantly reduced faecal egg counts, worm burdens and anaemia [79], warranting investigation of the protective potential of the *Haemonchus* APR-1-related protease. Other novel *Hc-apr* genes group phylogenetically with *C. elegans asp* genes (Figure S6 in Additional file 1) and are increased in the environmental L1 stage,

indicating a developmental role, consistent with *C. elegans* data [80, 81] and ruling these out as potential vaccine targets.

## Conclusions

*Haemonchus contortus* is the one of the most intensively researched parasitic nematode species and is the most established parasite species used in drug discovery, drug mode-of-action and drug resistance research. It is the first strongylid nematode to have its genome sequenced and the analysis presented here provides a first genomic insight into the biology of a gastro-intestinal strongylid and blood feeding parasitic nematode. It confirms the close relationship between *H. contortus* and *C. elegans* and provides a platform from which to apply the *C. elegans* biological knowledge to investigate the biology of this parasite for both basic and applied research. It also highlights specific differences between the species that will be important to inform researchers of those aspects of *C. elegans* biology that cannot be extrapolated. The genome sequence, gene models, transcriptome datasets and bioinformatic analyses provide a wealth of information on potential new drug and vaccine targets. They will also be an invaluable resource for the application of post-genomic technologies to this and other related strongylid parasites.

## Materials and Methods

### Parasite material production and quality assurance

The MHco3(ISE) line is a *H. contortus* strain, susceptible to all the major anthelmintic classes, that has been passaged in the laboratory by serial infection for many years. It has been extensively phenotypically and genetically characterised previously [82]. The inbred MHco3(ISE).N1 strain was derived by a genetically validated single pair mating of an adult male and an adult female MHco3(ISE) worm using a method of direct transplantation into the abomasum of a recipient parasite-free sheep (described in [83]). The inbred MHco3(ISE).N1 strain was subsequently passaged by oral experimental infection of parasite-free sheep and parasite material was harvested using standard methods (see Additional methods, Additional file 1 for details). All parasite material used in this study was derived from the inbred *H. contortus* MHco3(ISE).N1 strain except for the female gut transcriptomic data, which was generated from the MHco3(ISE) strain. The MHco3(ISE) and MHco3(ISE).N1 strains were routinely monitored for genetic integrity at each passage, as was all parasite material derived from the strains, using a standard panel of microsatellite markers and a well established “genetic fingerprinting” methodology [8, 82]. Both the MHco3(ISE) and the MHco3(ISE).N1 are viably cryopreserved and available on request.

### Genome sequencing and assembly

We assembled a draft sequence of the *H. contortus* genome based on data from a mixture of sequencing technologies, including 21x coverage of paired-end and shotgun reads from a Roche454SLX platform and approximately 163x coverage of long- and short-insert paired-end libraries from an Illumina HiSeq (coverage based on a genome size of 370Mb). Genomic and transcriptomic sequence data was generated using largely standard molecular biology methods (see Additional methods, Additional file 1), except that whole-genome amplified material was used to generate sufficient material for a large-insert Illumina library from a single male worm, using a



modified protocol. The genomic reads from each technology were initially assembled independently using assembly algorithms most suited to the typical coverage and read length of each, and scaffolds from each were then merged to form an initial set of scaffolds that were then improved by automated gap filling, followed by breaking and re-scaffolding using the paired-end data from both sequencing platforms.

### **Protein coding gene prediction and functional annotation**

After quality control and end-trimming, the transcriptome reads were mapped against the reference genome using TopHat software, v 2.0.6 (--mate-std-dev 50 -a 6 -i 10 -l 20000 --microexon-search --min-segment-intron 10 --max-segment-intron 50000) [84]. A reference dataset of 399 *H. contortus* protein encoding genes was manually curated from predictions of highly conserved genes using CEGMA (version 2.4) [24] and RNA-seq mapping. 347 of those were used to train Augustus [85] and 52 were used to independently evaluate the accuracy of the predictions. Final gene prediction (v2.0) was performed by Augustus using *H. contortus* specific parameters and RNA-seq, EST and polyA mappings as evidence hints generated by TopHat2 and PASA [86] respectively. Gene prediction accuracy was computed at the level of nucleotides, exons and complete genes on 57 manual curated gene models as described previously [87] and shown in table S7 (Additional file 1). Functional annotation information from the interpro databases using interproscan v4.5 [88], Gene Ontology (GO; [89]) terms were annotated via interpro2GO and from the curated *C. elegans* annotation in Wormbase (release 235;[90]) by assigning all GO terms shared by all *C. elegans* genes in a gene family to the *H. contortus* members of that family. Further functional insight was obtained by BLAST searches for similar genes in the GenBank nr database, and putative signal peptides were identified by SignalP [91]. To investigate *H. contortus* metabolism, a total of 828 enzyme classifications (ECs), covering 2,853 proteins, were assigned using KAAS [92], DETECT [93] and EFICAz [94]. Of these, 563 ECs covering 1,246 proteins were assigned to a metabolic pathway; the others are non-metabolic enzymes. Similar annotation efforts were carried out on *Caenorhabditis* species and *P. pacificus*.

## Gene expression and GO analysis

The numbers of RNA-seq reads per gene model were counted using custom-made scripts making use of BEDtools and a gff file of the genome annotation, and based on the TopHat mapping described above. Analysis of gene expression was performed using the DESeq (v 1.8.1) package for Bioconductor [95]. Read coverage was normalised to estimate the effective library sizes for each library and negative binomial tests performed between pairs of sample triplicates, using dispersion estimates from the default approach, to obtain p-values for differential expression of each gene adjusted for false discovery rate using the Benjamini-Hochberg procedure for multi-testing [96]. Only genes with adjusted p-values less or equal to  $1e^{-5}$  were retained. GO terms enriched in the set of differentially expressed genes in each comparison were identified using the “weight01” algorithm of the TopGO (v 2.8.0) package for Bioconductor [97]. Only GO terms with  $p < 0.01$  were considered for more detailed analysis. Expression data was drawn using Circos-0.62. All differentially expressed genes were clustered using MBCluster.seq with 50 clusters, and clusters then ordered based on which stage had the highest mean expression in that cluster.

## Data access

All sequences described in this paper have been submitted to the Genbank database, project ID PRJEB506. Sequence data is available at <http://www.ncbi.nlm.nih.gov/nuccore/CAVP000000000> and annotation has been submitted to Genbank and Wormbase.

The *H. contortus* genome assembly and functional annotation are available from <ftp://www.sanger.ac.uk/pub/pathogens/Haemonchus/contortus>.

## Acknowledgements

The *Haemonchus contortus* genome project is funded by the Wellcome Trust through their core support of the Wellcome Trust Sanger Institute (grant 098051) and a Wellcome Trust Project Grant to JSG (grant 067811). JSG also acknowledges funding from the Canadian Institutes of Health Research (CIHR) 230927. RL is funded through the Scottish Government by the Strategic Partnership for Animal Science Excellence (SPASE). SJ is funded through a BBSRC/KTN/Pfizer CASE studentship. We thank Nathalie Smerdon, Andrew Knights and David Willey for technical support in generating RNA-seq libraries; Coline Griffiths, Matthew Jones, Lesley Shirley, Naomi Park and Mike Quail for Illumina genomic libraries; and Richard Rance for 454 genomic libraries. Robert Davies and Quan Lin arranged for data release to the SRA. We thank Doug Jasmer for the *H. contortus* gut dissection protocol.

## Abbreviations

SL: spliced leader; CBL: Cathepsin B protease; GO: Gene Ontology; CYP: cytochrome P450; TCA: tricarboxylic acid; pLGIC: pentameric ligand-gated ion-channel; IVM: ivermectin; AChR: acetylcholine gated receptor; MPTL: monepantel; LEV: levamisole; SDR: short-chain dehydrogenase/reductase; UGT: UDP-glucuronosyl transferase; GST: glutathione S-transferase; ABC: ATP-binding cassette; APR: aspartic proteases; MEP: metalloproteases.

## Competing Interests

The authors declare that they have no competing interests.

## Author Contributions

MB, JAC, JSG and JDW designed research, which was coordinated by HB, MB, JSG and NH; DJB, ED, AAM and NS maintained the parasite life cycle, performed the inbreeding crosses and contributed biological material; RL, ER and GS prepared genomic DNA, RNA and transcriptome sequencing libraries; CB, RNB, DC, JAC, AG, MH, SJ, IK, TK, KL, RL, AM, AJR, IJT, AW and JDW analysed data; JAC,

JSG and RL drafted the complete manuscript, with sections of text contributed by CB, RNB, DC, AG, SJ, IK, TK, KL, AM, AJR, IJT, AW and JDW. All authors read and approved the final manuscript.

## References

1. Kaplan RM: **Drug resistance in nematodes of veterinary importance: a status report.** *Trends Parasitol* 2004, **20**:477-481.
2. Kaplan RM, Vidyashankar AN: **An inconvenient truth: global worming and anthelmintic resistance.** *Vet Parasitol* 2012, **186**:70-78.
3. Prichard RK, Basanez MG, Boatin BA, McCarthy JS, Garcia HH, Yang GJ, Sripa B, Lustigman S: **A research agenda for helminth diseases of humans: intervention for control and elimination.** *PLoS Negl Trop Dis* 2012, **6**:e1549.
4. Gilleard JS: **The use of *Caenorhabditis elegans* in parasitic nematode research.** *Parasitology* 2004, **128 Suppl 1**:S49-S70.
5. Geary TG, Thompson DP: ***Caenorhabditis elegans*: how good a model for veterinary parasites?** *Vet Parasitol* 2001, **101**:371-386.
6. Hashmi S, Tawe W, Lustigman S: ***Caenorhabditis elegans* and the study of gene function in parasites.** *Trends Parasitol* 2001, **17**:387-393.
7. Gilleard JS: **Understanding anthelmintic resistance: the need for genomics and genetics.** *Int J Parasitol* 2006, **36**:1227-1239.
8. Redman E, Sargison N, Whitelaw F, Jackson F, Morrison A, Bartley DJ, Gilleard JS: **Introgression of ivermectin resistance genes into a susceptible *Haemonchus contortus* strain by multiple backcrossing.** *PLoS Pathog* 2012, **8**:e1002534.

9. Kaminsky R, Ducray P, Jung M, Clover R, Rufener L, Bouvier J, Weber SS, Wenger A, Wieland-Berghausen S, Goebel T et al.: **A new class of anthelmintics effective against drug-resistant nematodes.** *Nature* 2008, **452**:176-180.
10. Taylor CM, Martin J, Rao RU, Powell K, Abubucker S, Mitreva M: **Using Existing Drugs as Leads for Broad Spectrum Anthelmintics Targeting Protein Kinases.** *PLoS Pathog* 2013, **9**:e1003149.
11. Knox DP, Redmond DL, Newlands GF, Skuce PJ, Pettit D, Smith WD: **The nature and prospects for gut membrane proteins as vaccine candidates for *Haemonchus contortus* and other ruminant trichostrongyloids.** *Int J Parasitol* 2003, **33**:1129-1137.
12. LeJambre LF, Windon RG, Smith WD: **Vaccination against *Haemonchus contortus*: performance of native parasite gut membrane glycoproteins in Merino lambs grazing contaminated pasture.** *Vet Parasitol* 2008, **153**:302-312.
13. Bethony JM, Loukas A, Hotez PJ, Knox DP: **Vaccines against blood-feeding nematodes of humans and livestock.** *Parasitology* 2006, **133 Suppl**:S63-S79.
14. Laing R, Hunt M, Protasio AV, Saunders G, Mungall K, Laing S, Jackson F, Quail M, Beech R, Berriman M et al.: **Annotation of two large contiguous regions from the *Haemonchus contortus* genome using RNA-seq and comparative analysis with *Caenorhabditis elegans*.** *PLoS One* 2011, **6**:e23216.
15. Blaxter ML, De LP, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM et al.: **A molecular evolutionary framework for the phylum Nematoda.** *Nature* 1998, **392**:71-75.
16. Taylor MA, Coop RL, Wall RL: *Veterinary Parasitology*. Oxford: Blackwell Publishing; 2007.

17. Prichard R: **Genetic variability following selection of *Haemonchus contortus* with anthelmintics.** *Trends Parasitol* 2001, **17**:445-453.
18. Waller PJ, Rudby-Martin L, Ljungstrom BL, Rydzik A: **The epidemiology of abomasal nematodes of sheep in Sweden, with particular reference to over-winter survival strategies.** *Vet Parasitol* 2004, **122**:207-220.
19. Hoberg EP, Lichtenfels JR, Gibbons L: **Phylogeny for species of *Haemonchus* (Nematoda: Trichostrongyloidea): Considerations of their evolutionary history and global biogeography among *Camelidae* and *Pecora* (Artiodactyla).** *Journal of Parasitology* 2004, **90**:1085-1102.
20. Blouin MS, Yowell CA, Courtney CH, Dame JB: **Host movement and the genetic structure of populations of parasitic nematodes.** *Genetics* 1995, **141**:1007-1014.
21. Troell K, Engstrom A, Morrison DA, Mattsson JG, Hoglund J: **Global patterns reveal strong population structure in *Haemonchus contortus*, a nematode parasite of domesticated ruminants.** *Int J Parasitol* 2006, **36**:1305-1316.
22. Gilleard JS, Beech RN: **Population genetics of anthelmintic resistance in parasitic nematodes.** *Parasitology* 2007, **134**:1133-1147.
23. Brasil BSAF, Nunes RL, Bastianetto E, Drummond MG, Carvalho DC, Leite RC, Molento MB, Oliveira DAA: **Genetic diversity patterns of *Haemonchus placei* and *Haemonchus contortus* populations isolated from domestic ruminants in Brazil.** *International Journal for Parasitology* 2012, **42**:469-479.
24. Leroy S, Duperray C, Morand S: **Flow cytometry for parasite nematode genome size measurement.** *Molecular and Biochemical Parasitology* 2003, **128**:91-93.

25. Parra G, Bradnam K, Korf I: **CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.** *Bioinformatics* 2007, **23**:1061-1067.
26. Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P et al.: **The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism.** *Nat Genet* 2008, **40**:1193-1198.
27. Scott AL, Ghedin E: **The genome of *Brugia malayi* - All worms are not created equal.** *Parasitology International* 2009, **58**:6-11.
28. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC et al.: **Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*.** *Nat Biotechnol* 2008, **26**:909-915.
29. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, Cohn J, Cromer J, Diener S, Gajan J, Graham S et al.: **Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism.** *Proc Natl Acad Sci U S A* 2008, **105**:14802-14807.
30. Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, McVeigh P, Takanashi T, Tsai IJ, Assefa SA, Cock PJ et al.: **Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*.** *PLoS Pathog* 2011, **7**:e1002219.
31. Guiliano DB, Hall N, Jones SJ, Clark LN, Corton CH, Barrell BG, Blaxter ML: **Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes.** *Genome Biol* 2002, **3**:RESEARCH0057.
32. Hoekstra R, Criado-Fornelio A, Fakkeldij J, Bergman J, Roos MH: **Microsatellites of the parasitic nematode *Haemonchus contortus*: polymorphism and linkage with a direct repeat.** *Molecular and Biochemical Parasitology* 1997, **89**:97-107.



33. Johnson PCD, Webster LMI, Adam A, Buckland R, Dawson DA, Keller LF: **Abundant variation in microsatellites of the parasitic nematode *Trichostrongylus tenuis* and linkage to a tandem repeat.** *Molecular and Biochemical Parasitology* 2006, **148**:210-218.
34. Grillo V, Jackson F, Gilleard JS: **Characterisation of *Teladorsagia circumcincta* microsatellites and their development as population genetic markers.** *Molecular and Biochemical Parasitology* 2006, **148**:181-189.
35. Allen MA, Hillier LW, Waterston RH, Blumenthal T: **A global analysis of *C. elegans* trans-splicing.** *Genome Research* 2011, **21**:255-264.
36. Rufener L, Maser P, Roditi I, Kaminsky R: ***Haemonchus contortus* Acetylcholine Receptors of the DEG-3 Subfamily and Their Role in Sensitivity to Monepantel.** *PLoS Pathog* 2009, **5**:e1000380.
37. Kotze AC: **Cytochrome P450 monooxygenase activity in *Haemonchus contortus* (Nematoda).** *Int J Parasitol* 1997, **27**:33-40.
38. Gerisch B, Antebi A: **Hormonal signals produced by DAF-9/cytochrome P450 regulate *C. elegans* dauer diapause in response to environmental cues.** *Development* 2004, **131**:1765-1776.
39. Laing ST, Ivens A, Butler V, Ravikumar SP, Laing R, Woods DJ, Gilleard JS: **The Transcriptional Response of *Caenorhabditis elegans* to Ivermectin Exposure Identifies Novel Genes Involved in the Response to Reduced Food Intake.** *PLoS ONE* 2012, **7**:e31367.
39. Fuchs S, Bundy J, Davies S, Viney J, Swire J, Leroi A: **A metabolic signature of long life in *Caenorhabditis elegans*.** *BMC Biology* 2010, **8**:14.

40. Weinstein PP: **Vitamin B12 changes in *Nippostrongylus brasiliensis* in its free-living and parasitic habitats with biochemical implications.** *Journal of Parasitology* 1996, **82**:1-6.
41. Yin Y, Martin J, Abubucker S, Scott AL, McCarter JP, Wilson RK, Jasmer DP, Mitreva M: **Intestinal Transcriptomes of Nematodes: Comparison of the Parasites *Ascaris suum* and *Haemonchus contortus* with the Free-living *Caenorhabditis elegans*.** *PLoS Negl Trop Dis* 2008, **2**:e269.
42. Chakrapani BP, Kumar S, Subramaniam JR: **Development and evaluation of an *in vivo* assay in *Caenorhabditis elegans* for screening of compounds for their effect on cytochrome P450 expression.** *J Biosci* 2008, **33**:269-277.
43. Lai CQ, Parnell LD, Lyman RF, Ordovas JM, Mackay TFC: **Candidate genes affecting *Drosophila* life span identified by integrating microarray gene expression analysis and QTL mapping.** *Mechanisms of Ageing and Development* 2007, **128**:237-249.
44. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB: **Computational Analysis of *Plasmodium falciparum* Metabolism: Organizing Genomic Information to Facilitate Drug Discovery.** *Genome Research* 2004, **14**:917-924.
45. Kushwaha S, Singh PK, Rana AK, Misra-Bhattacharya S: **Cloning, expression, purification and kinetics of trehalose-6-phosphate phosphatase of filarial parasite *Brugia malayi*.** *Acta Tropica* 2011, **119**:151-159.
46. Kantardjieff KA, Kim CY, Naranjo C, Waldo GS, Lakin T, Segelke BW, Zemla A, Park MS, Thomas C. ***Mycobacterium tuberculosis* RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway.** *Acta Crystallographica Section D Biological Crystallography* 60, 895-902. 2004.

47. Saunders GI, Wasmuth JD, Beech R, Laing R, Hunt M, Naghra H, Cotton JA, Berriman M, Britton C, Gilleard JS: **Characterization and comparative analysis of the complete *Haemonchus contortus*  $\beta$ -tubulin gene family and implications for benzimidazole resistance in strongylid nematodes.** *International Journal for Parasitology* 2013, **43**:465-475.
48. Forrester SG, Prichard RK, Dent JA, Beech RN: ***Haemonchus contortus*: HcGluCla expressed in *Xenopus* oocytes forms a glutamate-gated ion channel that is activated by ibotenate and the antiparasitic drug ivermectin.** *Molecular and Biochemical Parasitology* 2003, **129**:115-121.
49. Glendinning SK, Buckingham SD, Sattelle DB, Wonnacott S, Wolstenholme AJ: **Glutamate-Gated Chloride Channels of *Haemonchus contortus* Restore Drug Sensitivity to Ivermectin Resistant *Caenorhabditis elegans*.** *PLoS ONE* 2011, **6**:e22390.
50. Blackhall WJ, Pouliot JF, Prichard RK, Beech RN: ***Haemonchus contortus*: Selection at a Glutamate-Gated Chloride Channel Gene in Ivermectin- and Moxidectin-Selected Strains.** *Experimental Parasitology* 1998, **90**:42-48.
51. Beech R, Levitt N, Cambos M, Zhou S, Forrester SG: **Association of ion-channel genotype and macrocyclic lactone sensitivity traits in *Haemonchus contortus*.** *Molecular and Biochemical Parasitology* 2010, **171**:74-80.
52. Putrenko I, Zakikhani M, Dent JA: **A Family of Acetylcholine-gated Chloride Channel Subunits in *Caenorhabditis elegans*.** *Journal of Biological Chemistry* 2005, **280**:6392-6398.
53. De Graef J, Claerebout E, Vercruysse J, Wolstenholme A, Mitreva M, Geldhof P: **Gene expression and mutation analysis of the parasite-specific glutamate-gated chloride channel GLC-6 in resistant *Cooperia oncophora* isolates following *in vivo* exposure to macrocyclic lactones.** *International Journal for Parasitology: Drugs and Drug Resistance*, in press.

54. Jones A, Davis P, Hodgkin J, Sattelle D: **The nicotinic acetylcholine receptor gene family of the nematode *Caenorhabditis elegans*: an update on nomenclature.** *Invert Neurosci* 2007, **7**:129-131.
55. Liu Y, LeBeouf B, Guo X, Correa PA, Gualberto DG, Lints R, Garcia LR: **A Cholinergic-Regulated Circuit Coordinates the Maintenance and Bi-Stable States of a Sensory-Motor Behavior during *Caenorhabditis elegans* Male Copulation.** *PLoS Genet* 2011, **7**:e1001326.
56. Boulin T, Gielen M, Richmond JE, Williams DC, Paoletti P, Bessereau JL: **Eight genes are required for functional reconstitution of the *Caenorhabditis elegans* levamisole-sensitive acetylcholine receptor.** *Proceedings of the National Academy of Sciences* 2008, **105**:18590-18595.
57. Neveu C, Charvet CL, Fauvin A, Cortet J, Beech RN, Cabaret J: **Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance.** *Pharmacogenet Genomics* 2010, **20**:414-425.
58. Boulin T, Fauvin A, Charvet CL, Cortet J, Cabaret J, Bessereau JL, Neveu C: **Functional reconstitution of *Haemonchus contortus* acetylcholine receptors in *Xenopus* oocytes provides mechanistic insights into levamisole resistance.** *British Journal of Pharmacology* 2011, **164**:1421-1432.
59. Lindblom TH, Dodd AK: **Xenobiotic detoxification in the nematode *Caenorhabditis elegans*.** *J Exp Zool A Comp Exp Biol* 2006, **305**:720-730.
60. Kotze AC, Dobson RJ, Chandler D: **Synergism of rotenone by piperonyl butoxide in *Haemonchus contortus* and *Trichostrongylus colubriformis* in vitro: potential for drug-synergism through inhibition of nematode oxidative detoxification pathways.** *Vet Parasitol* 2006, **136**:275-282.

61. Jones BC, Middleton DS, Youdim K: **Cytochrome P450 Metabolism and Inhibition: Analysis for Drug Discovery.** In *Progress in Medicinal Chemistry*. Volume 47 edition. Edited by Edited by Lawton G and Witty DR. Amsterdam: Elsevier; 2009:239-263.
62. Sundaram P, Echalié B, Han W, Hull D, Timmons L: **ATP-binding Cassette Transporters Are Required for Efficient RNA Interference in *Caenorhabditis elegans*.** *Molecular Biology of the Cell* 2006, **17**:3678-3688.
63. Blackhall WJ, Prichard RK, Beech RN: **P-glycoprotein selection in strains of *Haemonchus contortus* resistant to benzimidazoles.** *Vet Parasitol* 2008, **152**:101-107.
64. Williamson SM, Storey B, Howell S, Harper KM, Kaplan RM, Wolstenholme AJ: **Candidate anthelmintic resistance-associated gene expression and sequence polymorphisms in a triple-resistant field isolate of *Haemonchus contortus*.** *Mol Biochem Parasitol* 2011, **180**:99-105.
65. Dicker AJ, Nisbet AJ, Skuce PJ: **Gene expression changes in a P-glycoprotein (Tci-pgp-9) putatively associated with ivermectin resistance in *Teladorsagia circumcincta*.** *Int J Parasitol* 2011, **41**:935-942.
66. Clark C., Kiesel G, Goby C: **Measurements of blood loss caused by *Haemonchus contortus* infection in sheep.** *Am J Vet Res* 1962, **23**:977-980.
67. Newton SE: **Progress on vaccination against *Haemonchus contortus*.** *International Journal for Parasitology* 1995, **25**:1281-1289.
68. Gluzman IY, Francis SE, Oksman A, Smith CE, Duffin KL, Goldberg DE: **Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway.** *J Clin Invest* 1994, **93**:1602-1608.

69. Williamson AL, Lecchi P, Turk BE, Choe Y, Hotez PJ, McKerrow JH, Cantley LC, Sajid M, Craik CS, Loukas A: **A Multi-enzyme Cascade of Hemoglobin Proteolysis in the Intestine of Blood-feeding Hookworms.** *Journal of Biological Chemistry* 2004, **279**:35950-35957.
70. Larminie CG, Johnstone IL: **Isolation and characterization of four developmentally regulated cathepsin B-like cysteine protease genes from the nematode *Caenorhabditis elegans*.** *DNA Cell Biol* 1996, **15**:75-82.
71. Ranjit N, Zhan B, Stenzel DJ, Mulvenna J, Fujiwara R, Hotez PJ, Loukas A: **A family of cathepsin B cysteine proteases expressed in the gut of the human hookworm, *Necator americanus*.** *Molecular and Biochemical Parasitology* 2008, **160**:90-99.
72. Pratt D, Armes LG, Hageman R, Reynolds V, Boisvenue RJ, Cox GN: **Cloning and sequence comparisons of four distinct cysteine proteases expressed by *Haemonchus contortus* adult worms.** *Molecular and Biochemical Parasitology* 1992, **51**:209-218.
73. Rehman A, Jasmer DP: **A tissue specific approach for analysis of membrane and secreted protein antigens from *Haemonchus contortus* gut and its application to diverse nematode species.** *Mol Biochem Parasitol* 1998, **97**:55-68.
74. Skuce PJ, Redmond DL, Liddel S, Stewart EM, Newlands GFJ, Smith W, Knox D: **Molecular cloning and characterization of gut-derived cysteine proteinases associated with a host protective extract from *Haemonchus contortus*.** *Parasitology* 1999, **119**:405-412.
75. Yatsuda AP, Bakker N, Krijgsveld J, Knox DP, Heck AJR, de Vries E: **Identification of Secreted Cysteine Proteases from the Parasitic Nematode *Haemonchus contortus* Detected by Biotinylated Inhibitors.** *Infection and Immunity* 2006, **74**:1989-1993.
76. Jasmer DP, Mitreva MD, McCarter JP: **mRNA sequences for *Haemonchus contortus* intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared**

- with other adult mammalian parasitic nematodes.** *Molecular and Biochemical Parasitology* 2004, **137**:297-305.
77. Smith SK, Smith WD: **Immunisation of sheep with an integral membrane glycoprotein complex of *Haemonchus contortus* and with its major polypeptide components.** *Res Vet Sci* 1996, **60**:1-6.
  78. Smith TS, Munn EA, Graham M, Tavernor AS, Greenwood CA: **Purification and evaluation of the integral membrane protein H11 as a protective antigen against *Haemonchus contortus*.** *International Journal for Parasitology* 1993, **23**:271-280.
  79. Loukas A, Bethony JM, Mendez S, Fujiwara RT, Goud GN, Ranjit N, Zhan B, Jones K, Bottazzi ME, Hotez PJ: **Vaccination with Recombinant Aspartic Hemoglobinase Reduces Parasite Load and Blood Loss after Hookworm Infection in Dogs.** *PLoS Med* 2005, **2**:e295.
  80. Roy PJ, Stuart JM, Lund J, Kim SK: **Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*.** *Nature* 2002, **418**:975-979.
  81. Tcherepanova I, Bhattacharyya L, Rubin CS, Freedman JH: **Aspartic proteases from the nematode *Caenorhabditis elegans*. Structural organization and developmental and cell-specific expression of asp-1.** *J Biol Chem* 2000, **275**:26359-26369.
  82. Redman E, Packard E, Grillo V, Smith J, Jackson F, Gilleard JS: **Microsatellite analysis reveals marked genetic differentiation between *Haemonchus contortus* laboratory isolates and provides a rapid system of genetic fingerprinting.** *Int J Parasitol* 2008, **38**:111-122.
  83. Sargison N: **Development of genetic crossing methods to identify genes associated with macrocyclic lactone resistance in the sheep nematode parasite, *Haemonchus contortus*.** *PhD Thesis*. University of Edinburgh; 2009.

84. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**:1105-1111.
85. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B: **AUGUSTUS: ab initio prediction of alternative transcripts.** *Nucleic Acids Research* 2006, **34**:W435-W439.
86. Haas B, Salzberg S, Zhu W, Pertea M, Allen J, Orvis J, White O, Buell CR, Wortman J: **Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments.** *Genome Biology* 2008, **9**:R7.
87. Guigo R, Flicek P, Abril J, Reymond A, Lagarde J, Denoeud F, Antonarakis S, Ashburner M, Bajic V, Birney E et al.: **EGASP: the human ENCODE Genome Annotation Assessment Project.** *Genome Biology* 2006, **7**:S2.
88. Zdobnov EM, Apweiler R: **InterProScan - an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847-848.
89. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al.: **Gene Ontology: tool for the unification of biology.** *Nat Genet* 2000, **25**:25-29.
90. Yook K, Harris TW, Bieri T, Cabunoc A, Chan J, Chen WJ, Davis P, De La Cruz N, Duong A, Fang R et al.: **WormBase 2012: more genomes, more data, new website.** *Nucleic Acids Research* 2012, **40**:D735-D741.
91. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Meth* 2011, **8**:785-786.
92. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and pathway reconstruction server.** *Nucleic Acids Research* 2007, **35**:W182-W185.



93. Hung SS, Wasmuth J, Sanford C, Parkinson J: **DETECT - A Density Estimation Tool for Enzyme Classification and its application to Plasmodium falciparum.** *Bioinformatics* 2010, **26**:1690-1698.
94. Arakaki A, Huang Y, Skolnick J: **EFICAz2: enzyme function inference by a combined approach enhanced by machine learning.** *BMC Bioinformatics* 2009, **10**:107.
95. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biology* 2010, **11**:R106.
96. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995, **57**:289-300.
97. Alexa A, Rahnenfuhrer J. **topGO: Enrichment analysis for Gene Ontology, R package version 2.8.0.** 2010.

## Figure Legends

Figure 1. Evolution of genome content in Clade V nematodes. A maximum-likelihood phylogeny based on concatenated alignment of single-copy genes. Values on edges represent the inferred numbers of births (+) and deaths (-) of gene families along that edge. Note that our approach cannot distinguish gene family losses from gains on the basal branches of this tree, so for example the value of 1263 gene family gains on the basal branch of clade V will include gene families lost on the branch leading to the clade IV spp. The first column of pie charts represent the gene family composition of each genome – the area of the circle is proportional to the predicted proteome size, and wedges represent the numbers of proteins predicted to be either singletons (i.e. not members of any gene family), members of gene families common to all nine genomes, members of gene families present only in a single genome, and members of all other gene families. The second column of pie charts represents the genomic composition of the species with published genome sequences. All ingroup nodes had 100% bootstrap support.

Figure 2. One-to-one orthologs connecting *H. contortus* scaffolds and *C. elegans* chromosomes. Shown are (A) all 130 *H. contortus* scaffolds with at least five one-to-one orthologs between the two species, with scaffolds ordered to maximise colinearity with the *C. elegans* genome. (B) shows the relative positions of one-to-one orthologs between the two largest *H. contortus* scaffolds and *C. elegans* chromosomes, demonstrating the lack of conservation of gene order within syntenic blocks.

Figure 3. Differential gene expression across the *H. contortus* lifecycle. (A) Numbers of genes significantly up-regulated (red) and down-regulated (blue) between life-stages. (B) The expression profiles of genes determined to be differentially expressed between any pair of stages are shown in clusters ordered by the stage at which they are most highly expressed. The innermost circle shows normalized expression levels during the egg stage with further concentric circles showing expression

in the L1, L3, L4 stages, followed by male worms, female worms and then female gut samples. Mean expression values were taken over three replicates and are shown on a log2 scale.

Figure 4. Maximum-likelihood phylogenies of *H. contortus* and *C. elegans* ligand-gated ion channel families. Figure shows unrooted trees for (A) cation channel and (B) anion channel genes from the two nematode species, with different colours representing each major clade of channels. Within clades, *C. elegans* genes are coloured a lighter shade than *H. contortus* genes. Both trees show a general pattern of conserved one-to-one orthology, and coloured bars indicate where this pattern is broken – by a duplication in *H. contortus* (with four copies of *unc-29*), and by loss of some genes – there are four *H. contortus* genes (*acr-26*, *acr-27*, *glc-5*, *glc-6*) without orthologs in *C. elegans*, and 7 in *C. elegans* missing from *H. contortus* (*acr-13/lev-8*, *acr-18*, *acr-23*, *glc-1*, *lev-8*, *lgc-48*, *lgc-54*). All of these genes are shown in large bold type. A truncated ortholog of *C. elegans* *acr-9* is present in the *H. contortus* but is omitted from this phylogeny. Values near internal nodes are SH-test support values – with support values of 1.0 on all other nodes omitted for clarity.

Figure 5. (A) Maximum-likelihood phylogeny of cathepsin B (*cbl*) genes in *H. contortus* and related nematodes. The *H. contortus* genome shows large, independent expansions of all three main classes of *cbl* genes: AC (anticoagulant)-like, GCP-7-like and Hmcp-like. (B) Many of the genes in two of these expansions occur as tandem arrays of duplicates on three scaffolds of the *H. contortus* genome. Two of these scaffolds are known to be linked, as eight of the genes occur on a previously completed BAC sequence.

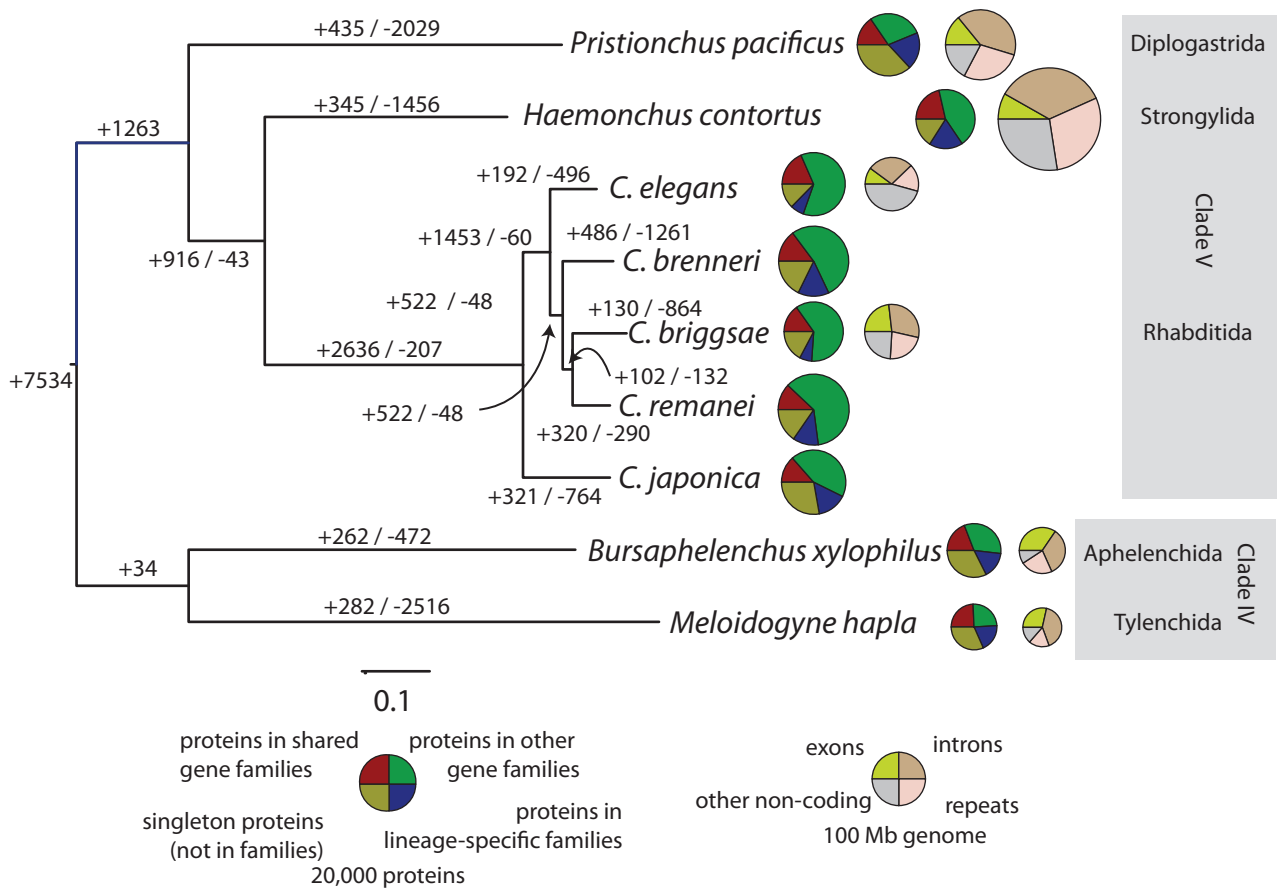


Figure 1

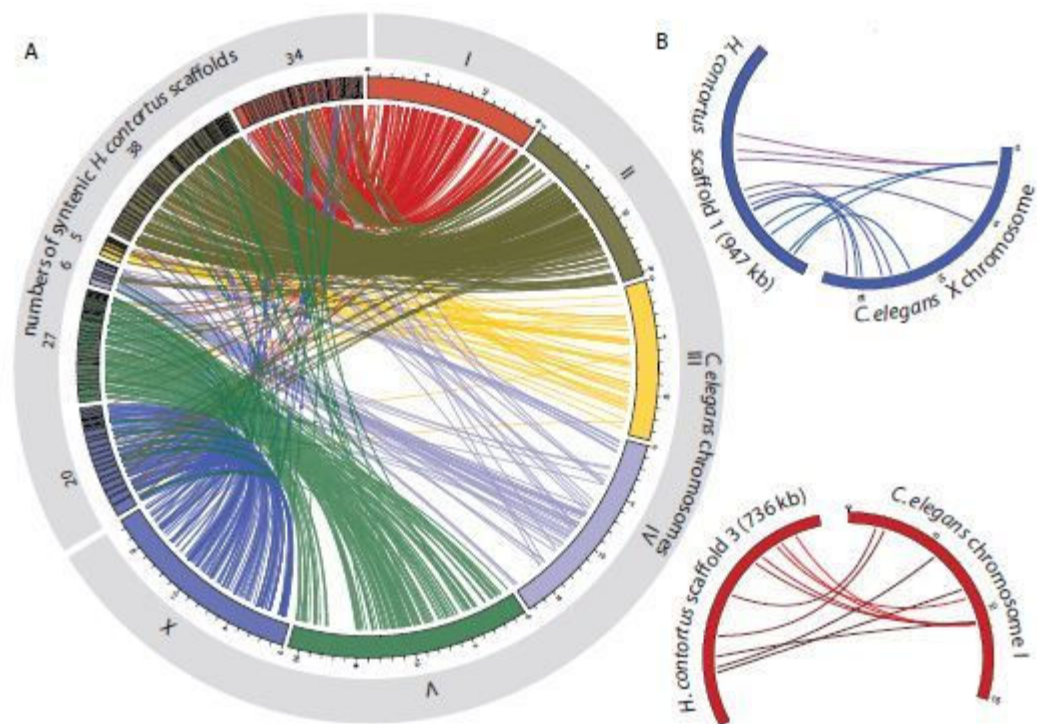
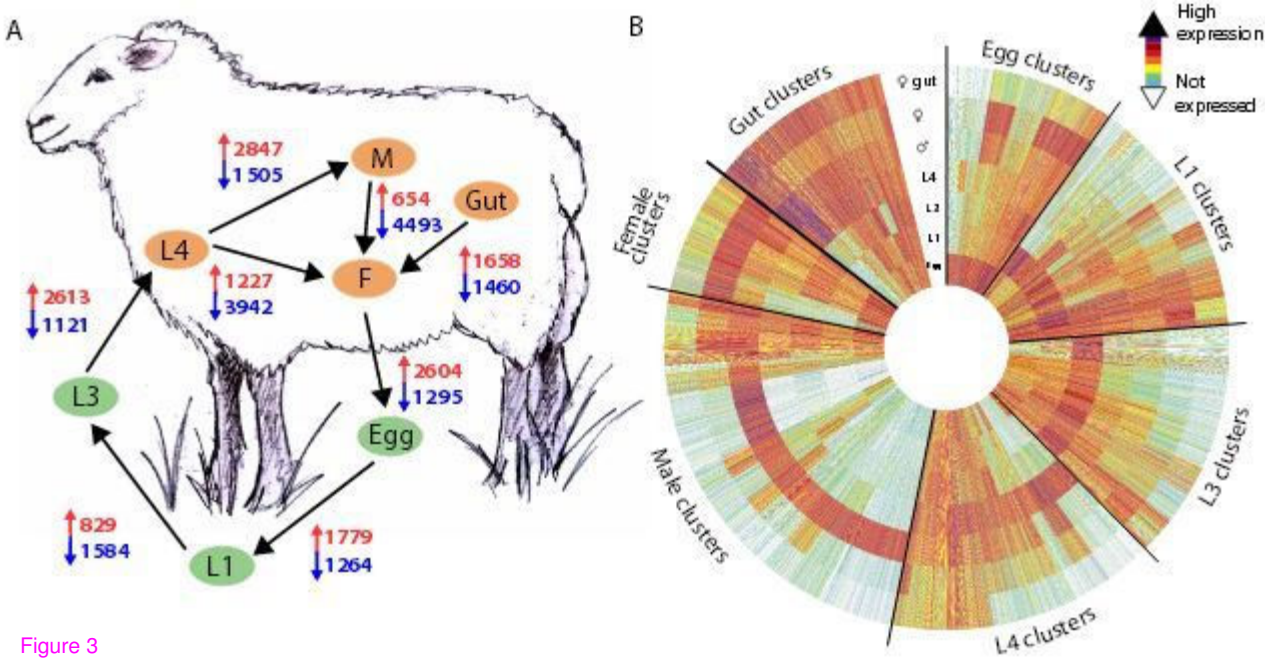


Figure 2



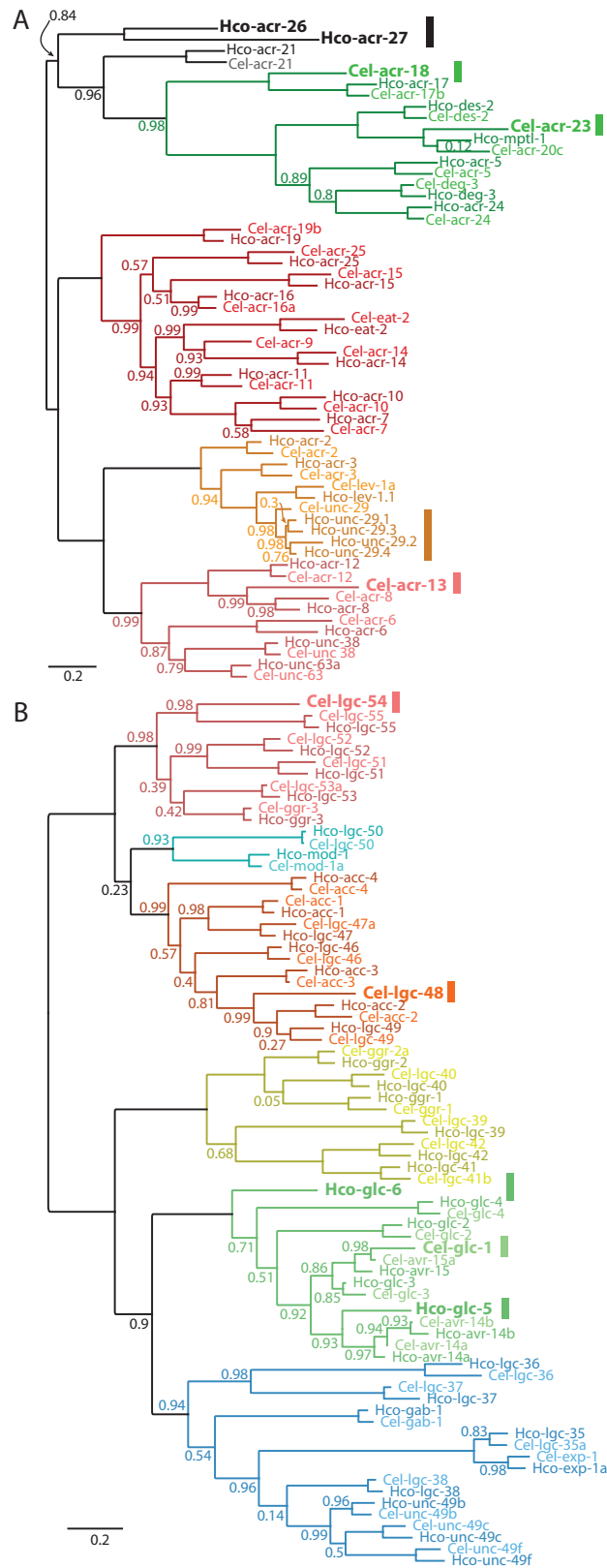


Figure 4

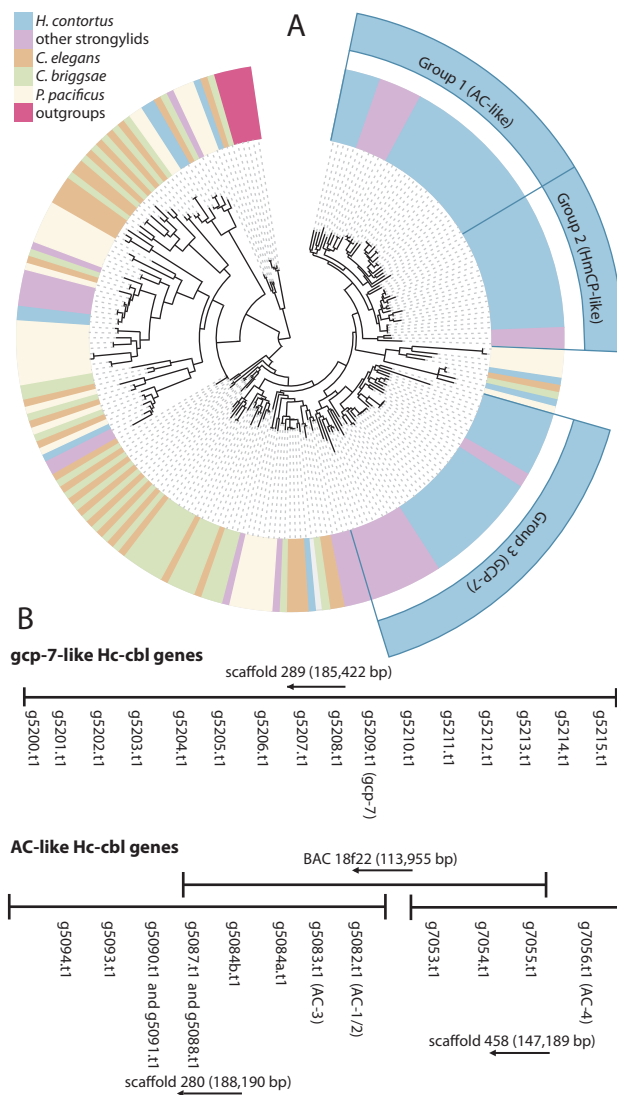


Figure 5



**Additional files provided with this submission:**

Additional file 1: Hc\_Appendix\_revision\_final.doc, 4122K

<http://genomebiology.com/imedia/1515792380102025/supp1.doc>

Additional file 2: Hc\_Genome\_Biology\_revision\_trackchanges.doc, 157K

<http://genomebiology.com/imedia/1668611319102025/supp2.doc>